AEM: PS5

Maxwell Austensen

December 19, 2016

```
library(tidyverse)
library(stringr)
library(lubridate)
library(haven)
library(stargazer)
library(sandwich)
library(plm)
library(lmtest)

# HTML or LaTeX rendering
type <- "latex"
```

Minimum Wage

```
minwage <- read_stata("http://users.nber.org/~rdehejia/!@$AEM/Problem%20Sets/ps5/DinD_ex.dta")

glimpse(minwage)
```

```
## Observations: 698
## Variables: 6
## $ sheet   <dbl> 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 9, 9, 10, 10, 11, ...
## $ fte     <dbl> 31.00, 40.00, 13.00, 12.50, 12.50, 7.50, 16.00, 20.00,...
## $ nj      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ after   <dbl> 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, ...
## $ njafter <dbl> 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, ...
## $ dfte    <dbl> NaN, 9.00, NaN, -0.50, NaN, -5.00, NaN, 4.00, NaN, 5.0...
```

1

```
diffs <-
  minwage %>%
  group_by(nj, after) %>%
  summarise(emp_mean = mean(fte)) %>%
  spread(after, emp_mean) %>%
  mutate(diff = `1` - `0`)

diffs$diff[[2]] - diffs$diff[[1]]
```

```
## [1] 2.328724
```

2

```r
# Helper function to fit model and get robust standard errors
model_prep <- function(df, f, robust = FALSE){
  mod <- lm(f, data = df)

  if(robust == TRUE){
    # Robust stanadard errors (replicating Stat's robust option)
    robust_se <-
      mod %>%
      vcovHC(type = "HC1") %>%
      diag() %>%
      sqrt()

    return(list(mod, robust_se))
  } else {
    return(list(mod))
  }
}
```

```r
m1 <- minwage %>% model_prep("dfte ~ nj")

stargazer(m1[1], type = type, header = FALSE)
```

Table 1:

	Dependent variable:
	dfte
nj	2.329**
	(1.178)
Constant	−2.046*
	(1.063)
Observations	349
R^2	0.011
Adjusted R^2	0.008
Residual Std. Error	8.570 (df = 347)
F Statistic	3.905** (df = 1; 347)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
m2 <- minwage %>% model_prep("dfte ~ nj", robust = TRUE)

stargazer(m2[1], se = m2[2], type = type, header = FALSE)
```

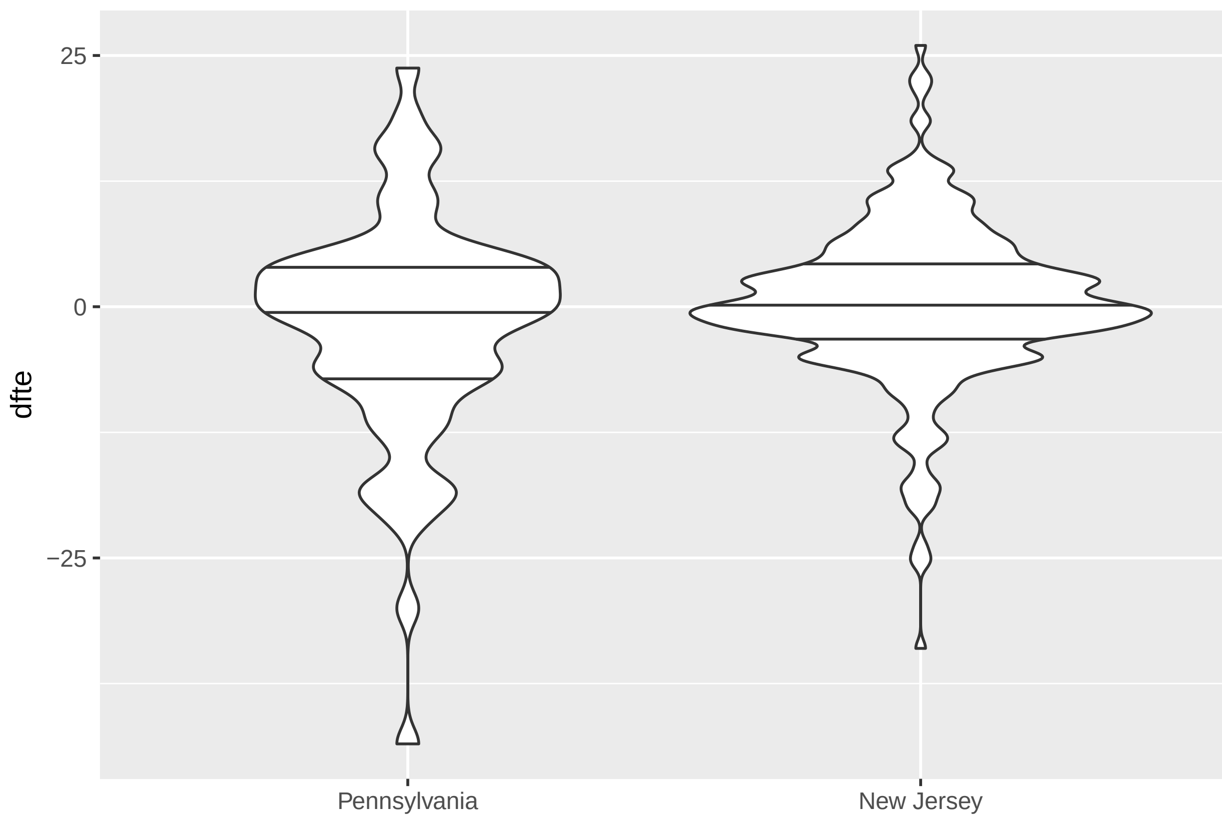
The difference in the standard errors is sizable because there is a considerable amount of heteroskedasticity to adjust for in the outcome variable `dfte`.

```r
minwage %>%
  ggplot(aes(factor(nj), dfte)) +
  geom_violin(adjust = .5, draw_quantiles = c(0.25, 0.5, 0.75)) +
  scale_x_discrete(labels = c("Pennsylvania", "New Jersey")) +
```

Table 2:

	Dependent variable:
	dfte
nj	2.329
	(1.470)
Constant	−2.046
	(1.395)
Observations	349
R^2	0.011
Adjusted R^2	0.008
Residual Std. Error	8.570 (df = 347)
F Statistic	3.905** (df = 1; 347)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
labs(x ="")
```



3

```r
m3 <- minwage %>% model_prep("fte ~ nj + after + njafter")

stargazer(m3[1], type = type, header = FALSE)
```

Table 3:

	Dependent variable:
	fte
nj	-2.999^{**}
	(1.233)
after	-2.046
	(1.572)
njafter	2.329
	(1.743)
Constant	20.300^{***}
	(1.112)
Observations	698
R^2	0.009
Adjusted R^2	0.005
Residual Std. Error	8.964 (df = 694)
F Statistic	2.088 (df = 3; 694)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
m4 <- minwage %>% model_prep("fte ~ nj + after + njafter", robust = TRUE)

stargazer(m4[1], se = m4[2], type = type, header = FALSE)
```

The coefficient of interest in these models is the interaction term `njafter` for NJ (treatment) and after (post-intervention), which takes the value 2.3287243. The robust standard errors are considerably larger, but in both cases this interaction term is not statistically significant at the 10% level.

4

```r
m5 <- plm(fte ~ nj + after + njafter, data = minwage, model = "pooling", index = c("sheet"))

# calculate small sample size adjustment
G <- length(unique(minwage$sheet))
N <- length(minwage$sheet)
dfa <- (G/(G - 1)) * (N - 1)/m5$df.residual

# use coeftest and the vcovHC functions, specifying HC0 type and identifying cluster as 'group'
m5_clus <- coeftest(m5, vcov = function(x) dfa * vcovHC(x, cluster="group", type="HC0"))

stargazer(m5_clus, type = type, dep.var.labels = "fte", header = FALSE)
```

Table 4:

	Dependent variable:
	fte
nj	-2.999^{*}
	(1.591)
after	-2.046
	(1.789)
njafter	2.329
	(1.931)
Constant	20.300^{***}
	(1.502)
Observations	698
R^2	0.009
Adjusted R^2	0.005
Residual Std. Error	8.964 (df = 694)
F Statistic	2.088 (df = 3; 694)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5:

	Dependent variable:
	fte
nj	-2.999^{*}
	(1.593)
after	-2.046
	(1.396)
njafter	2.329
	(1.471)
Constant	20.300^{***}
	(1.503)
Note:	*p<0.1; **p<0.05; ***p<0.01

When clustering the standard errors by restaurant they become much larger. They are not as large as the heteroskedasticity robust standard errors from the earlier model for the `after` and `njafter` terms, but as large for `nj`.

5

```
m6 <- plm(fte ~ nj + after + njafter, data = minwage, model = "within", index = c("sheet"))

stargazer(m6, type = type, header = FALSE)
```

Table 6:

	Dependent variable:
	fte
after	−2.046*
	(1.063)
njafter	2.329**
	(1.178)
Observations	698
R^2	0.011
Adjusted R^2	−0.986
F Statistic	2.007 (df = 2; 347)
Note:	*p<0.1; **p<0.05; ***p<0.01

When using restaurant fixed effects the `nj` term drops out because the state that a restaurant is in is invariant within restaurants over time.

6

The estimated impact of the minimum wage is the same in all these models because they are all essentially doing the same thing, taking the difference in differences, in different ways. The first models take the within group time differences in the outcome variable, and then include the group differences in the treatment term. The second set of models includes both the time and group terms and gets the difference with the interaction term.

MicroFinance

```
safesave <- read_stata("http://users.nber.org/~rdehejia/!@$AEM/Problem%20Sets/ps5/safesave_slim_data.dta

glimpse(safesave)
```

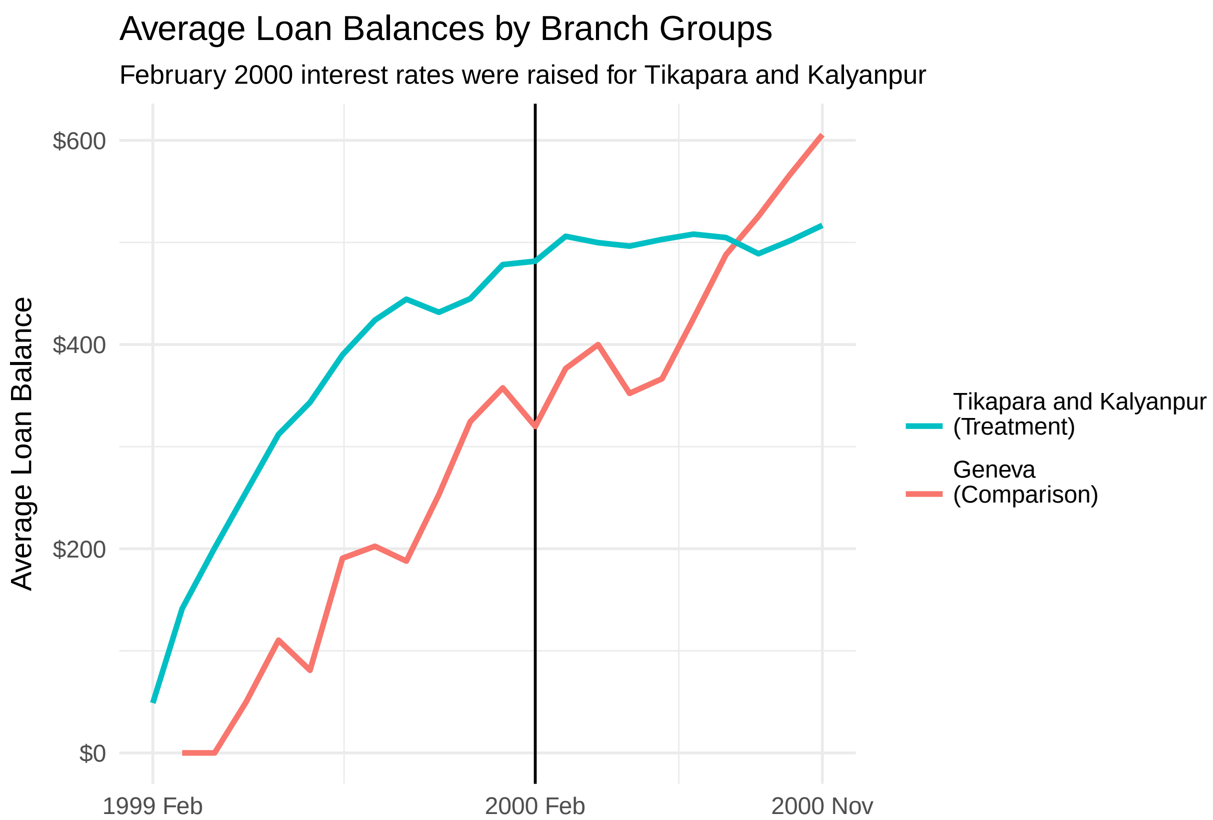
```
## Observations: 60,506
## Variables: 12
```

```
## $ nacc      <dbl> 202195, 101611, 208182, 202017, 208150, 208248, 1022...
## $ monthyear <dbl> 199902, 199902, 199902, 199902, 199902, 199902, 1999...
## $ saved     <dbl> 0.0000000, 8.9601719, 20.1603869, 0.0000000, 6.72012...
## $ withdrawn <dbl> 0.000000, 0.000000, 2.240043, 0.000000, 0.000000, 0....
## $ nage      <dbl> 23, 26, 13, 13, 31, 19, 31, 61, 8, 25, 36, 41, 26, 2...
## $ tinpr     <dbl> 11, 5, 7, 16, 9, 5, 4, 3, 13, 12, 7, 13, 7, 3, 3, 9,...
## $ loanbal   <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.00...
## $ TI        <dbl> 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0...
## $ KA        <dbl> 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1...
## $ GE        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ TIKA      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ trend     <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
```

```r
post_start <- as_date("2000-02-01")

plot_data <-
  safesave %>%
  mutate(year = monthyear %>% str_sub(1, 4) %>% as.numeric,
         month = monthyear %>% str_sub(5, 6) %>% as.numeric,
         date = make_date(year, month),
         treatment = recode(TIKA, `1` = "Tikapara and Kalyanpur \n(Treatment)\n",
                                  `0` = "Geneva \n(Comparison)\n"),
         post = if_else(date >= post_start, 1, 0),
         treat_and_post = str_c(treatment, post),
         days = date - min(date)) %>%
  group_by(TIKA, treatment, date, days, post, treat_and_post) %>%
  summarise(loan_balance = mean(loanbal),
            nage = mean(nage),
            tinpr = mean(tinpr)) %>%
  ungroup()
```

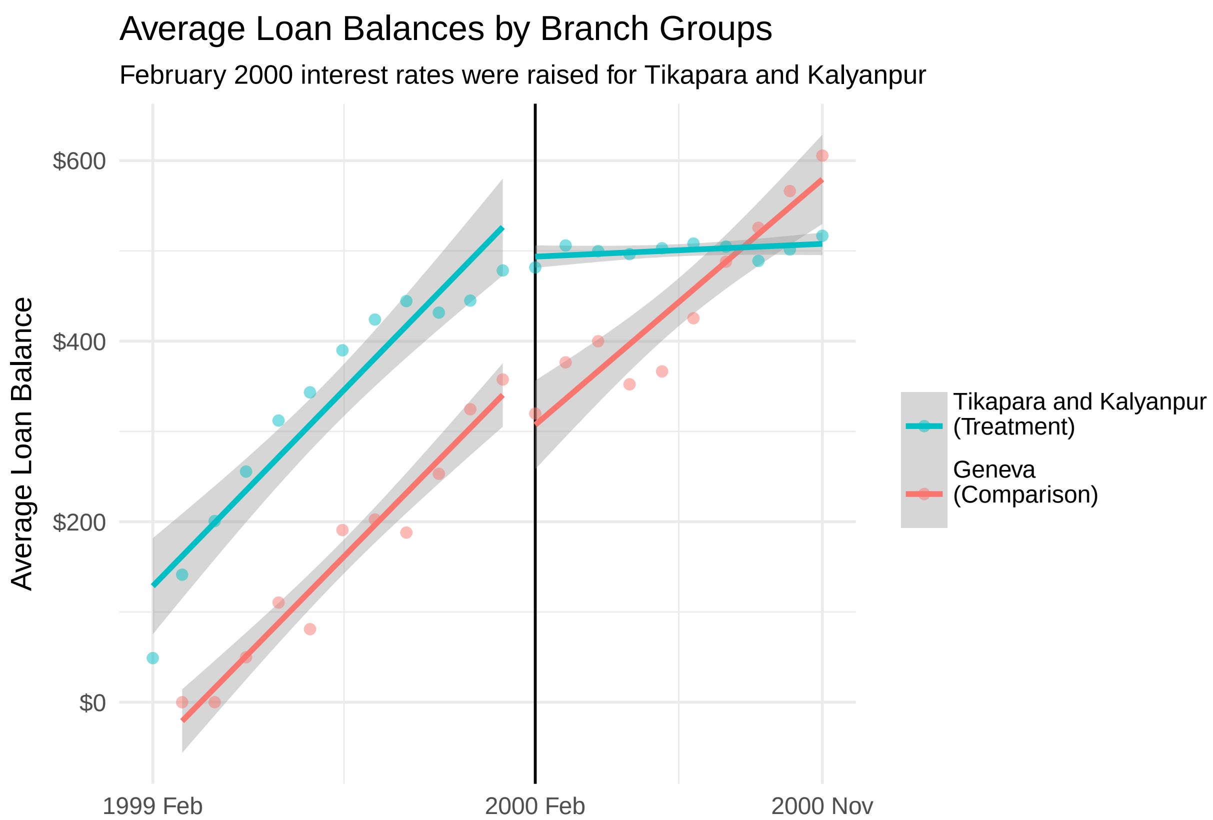
```r
start_date <- min(plot_data$date)
end_date <- max(plot_data$date)

base_plot <-
  plot_data %>%
  ggplot(aes(date, loan_balance, group = treatment, color = treatment)) +
  geom_vline(xintercept = as.numeric(post_start)) +
  scale_y_continuous(labels = scales::dollar) +
  scale_x_date(breaks = c(start_date, post_start, end_date), date_labels = "%Y %b") +
  guides(color = guide_legend(reverse = TRUE)) +
  theme_minimal() +
  labs(title = "Average Loan Balances by Branch Groups",
       subtitle = "February 2000 interest rates were raised for Tikapara and Kalyanpur",
       color = "",
       x = "",
       y = "Average Loan Balance")
```

```r
base_plot + geom_line(size = 1)
```



Eyeballing the figure, the pre-treatment tends look roughly parallel.

```
base_plot +
  geom_smooth(aes(date, loan_balance, group = treat_and_post, color = treatment),
              method = "lm", formula = y ~ x, size = 1) +
  geom_point(alpha = 0.5)
```




```
plot_data %>%
  filter(post == 0) %>%
  with(lm(loan_balance ~ days + TIKA + days*TIKA)) %>%
  stargazer(type = type, header = FALSE)
```

In the regression, the fact that the interaction term of the time trend and treatment `days:TIKA` is not statistically significant indicates that the treatment and comparison groups do not have significantly different pre-treatment time trends.

```
plot_data %>%
  filter(TIKA == 0) %>%
  with(lm(loan_balance ~ days + post + days*post)) %>%
  stargazer(type = type, header = FALSE)
```

Within the comparison branches there is no statistically significant change in the slope after the intervention (because the coefficient on the `days:post` interaction term is not significant), and neither does the intercept change significantly (because the `post` coefficient is also not significant).

```
plot_data %>%
  filter(post == 0) %>%
  with(lm(loan_balance ~ days + TIKA + days*TIKA + nage + tinpr)) %>%
  stargazer(type = type, header = FALSE)
```

It is not a problem for the validity of the identification strategy that the treatment and comparison groups have different pre-treatment levels, because by taking the difference in differences this is netted out. The issue would arise id the pre-treatment slopes where different. Also, the differences in pre-treatment levels is no longer statistically significant when controlling for the average age or borrowers and average length of time the borrower has been with the bank, as indicated by the not significant `TIKA` term in the above model.

Table 7:

	Dependent variable:
	loan_balance
days	1.181***
	(0.116)
TIKA	182.576***
	(31.241)
days:TIKA	0.010
	(0.155)
Constant	−53.988**
	(23.894)
Observations	23
R^2	0.949
Adjusted R^2	0.941
Residual Std. Error	37.345 (df = 19)
F Statistic	117.380*** (df = 3; 19)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8:

	Dependent variable:
	loan_balance
days	1.181***
	(0.100)
post	−1.143
	(62.368)
days:post	−0.188
	(0.153)
Constant	−53.988**
	(20.523)
Observations	21
R^2	0.973
Adjusted R^2	0.969
Residual Std. Error	32.077 (df = 17)
F Statistic	206.754*** (df = 3; 17)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9:

	Dependent variable:
	loan_balance
days	1.453***
	(0.386)
TIKA	393.425
	(243.242)
nage	−2.488
	(14.532)
tinpr	−26.460
	(27.184)
days:TIKA	0.181
	(0.278)
Constant	21.349
	(456.638)
Observations	23
R^2	0.952
Adjusted R^2	0.937
Residual Std. Error	38.385 (df = 17)
F Statistic	66.862*** (df = 5; 17)
Note:	*p<0.1; **p<0.05; ***p<0.01